# PLOS ONE

# Admixture mapping identifies genetic regions associated with blood pressure phenotypes in African Americans

Zhi Liu[1], Daniel Shriner[2], Nancy F. Hansen[1], Charles N. Rotimi[2], James C. Mullikin[1,3]*, on behalf of the NISC Comparative Sequencing Program[3]¶

1 Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, 2 Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, 3 NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America

¶ Author contributors of the NISC Comparative Sequencing Program are listed in the Acknowledgment section.
* mullikin@mail.nih.gov

## Abstract

Hypertension occurs at a higher rate in African Americans than in European Americans. Based on the assumption that causal variants are more frequently found on DNA segments inherited from the ancestral population with higher disease risk, we employed admixture mapping to identify genetic loci with excess local African ancestry associated with blood pressure. Chromosomal regions 1q21.2–21.3, 4p15.1, 19q12 and 20p13 were significantly associated with diastolic blood pressure (β = 5.28, -7.94, -6.82 and 5.89, $P$-value = 6.39E-04, 2.07E-04, 6.56E-05 and 5.04E-04, respectively); 1q21.2–21.3 and 19q12 were also significantly associated with mean arterial pressure (β = 5.86 and -6.40, $P$-value = 5.32E-04 and 6.37E-04, respectively). We further selected SNPs that had large allele frequency differences within these regions and tested their association with blood pressure. SNP rs4815428 was significantly associated with diastolic blood pressure after Bonferroni correction (β = -2.42, $P$-value = 9.57E-04), and it partially explained the admixture mapping signal at 20p13. SNPs rs771205 (β = -1.99, $P$-value = 3.37E-03), rs3126067, rs2184953 and rs58001094 (the latter three exhibit strong linkage disequilibrium, β = -2.3, $P$-value = 1.4E-03) were identified to be significantly associated with mean arterial pressure, and together they fully explained the admixture signal at 1q21.2–21.3. Although no SNP at 4p15.1 showed large ancestral allele frequency differences in our dataset, we detected association at low-frequency African-specific variants that mapped predominantly to the gene *PCDH7*, which is most highly expressed in aorta. Our results suggest that these regions may harbor genetic variants that contribute to the different prevalence of hypertension.

# Introduction

Hypertension is a strong risk factor for coronary artery disease. Hypertension is a heritable disease, with an estimated heritability of 30–50% [1]. So far, genome-wide association studies (GWAS) have been successful at identifying genetic loci associated with hypertension, but the effect sizes of individual loci have been small, and currently-identified loci only explain a small fraction of the total heritability (4–6%) [2]. Other genetic determinants for blood pressure are still yet to be discovered.

One way to investigate the unexplained heritability is to study less represented populations, such as African Americans. In the United States, hypertension is diagnosed more frequently in African Americans than European Americans, representing 40% of African Americans vs. 28% of European Americans among people from all age groups [3], and 60% of African Americans vs. 38% of European Americans among individuals aged between 45 to 85 years [4]. African Americans form an admixed population in which each individual not only displays different proportions of European and African ancestry, but their chromosomes also show ancestral mosaicisms resulting from recombination across generations. To identify causal genetic loci in admixed populations, admixture mapping serves as a powerful tool when large allele frequency differences are present in the ancestral populations. Ancestral allele frequency differences at causal loci may contribute to disease prevalence differences in different populations.

Previously, GWAS for hypertension have identified genetic loci such as those containing mutations in genes *PDE3A* [5], *NOS3* [6] and *CYP17A1-CNNM2-NT5C2* [7]. In addition to GWAS, admixture mapping has also been successful at identifying ancestral haplotypes significantly associated with hypertension. Up until now, several loci have been reported to be associated with blood pressure, such as 6q24 and 21q21 [8]; within 21q21, *CXADR* was likely to play a role in blood pressure in African Americans [9]; by utilizing CARe consortium data, 5p13 was identified to be associated with diastolic blood pressure (DBP), with 3 uncorrelated SNPs within this region adequately accounting for the observed association [10]. For Hispanics, 6p12.3 was found to be associated with local African ancestry for mean arterial pressure (MAP) and DBP, but no variants were identified that drove these associations [11]. Compared to GWAS, admixture mapping has a much lower testing burden and thus requires much smaller sample sizes.

In this study, we utilized the ClinSeq® cohort exome sequencing data to identify genetic regions where local African ancestry was associated with blood pressure phenotypes. We fine-mapped these regions and identified genetic variants with large ancestral allele frequency differences that drove local ancestral associations. Furthermore, we replicated these variants identified from ClinSeq® in an independent cohort of Africans.

# Materials and methods

## Patient samples and exome-sequencing

The ClinSeq® study was approved by the Institutional Review Boards at the National Institutes of Health and informed consent was obtained from each participant. The ClinSeq® A2 cohort, which consists of 503 unrelated participants of self-reported African descent, aged between 45–65, were recruited between 2012–2017 and seen at the Clinical Center of the National Institutes of Health. Participants were not ascertained based on any particular phenotype and were interviewed and measured for various anthropometric and clinical variables. Blood samples were collected, from which DNA was isolated and exomic regions of interest were captured using the Integrated DNA Technologies (IDT) capture kit. Whole exome sequencing was

performed at the NIH Intramural Sequencing Center, Rockville, MD. Variant calling was performed using bam2mpg [12]. As a quality control (QC) step, single nucleotide polymorphisms (SNPs) were filtered for $GQ \geq 10$ and $GQ/DP > 0.5$; only autosomal SNPs were retained; SNPs discovered to be out of Hardy-Weinberg equilibrium (*P*-value < 5.7E-07) based on an exact test [13], monomorphic SNPs, SNPs with call rate < 0.98, and SNPs with minor allele frequency (MAF) < 0.005 were removed. In a subsequent QC step, samples with mismatched sex, samples from related individuals and samples without phenotypes were removed. Based on principal component (PC) plots, samples outside of ± 4 standard deviations from the African American cluster by the first two eigenvalues were removed.

### Phenotypes

DBP and systolic blood pressure (SBP) were measured using a Dinamap instrument and obtained from the left arm after five minutes of rest with the subject in a sitting position with legs uncrossed. Two of these measurements were obtained and an average was taken for our analysis. For patients who took anti-hypertension medication, 5 mmHg and 10 mmHg were added to DBP and SBP, respectively [14]. MAP values were calculated as 1/3 * medically adjusted SBP + 2/3 * medically adjusted DBP. To check that values were normally distributed, we used a Shapiro-Wilk test. Untransformed DBP ($W = 0.99528$, *P*-value = 0.1424), MAP ($W = 0.9942$, *P*-value = 0.05881) and $Log_{10}$-transformed SBP ($W = 0.996$, *P*-value = 0.2496) were found to be normally distributed (S1 Fig).

### Admixture mapping

Local ancestry was inferred using SEQMIX v0.12 [15]. For ancestry inference purposes, SNPs were pruned differently than in the association study. All markers, regardless of MAF, were retained. To ensure SNPs that were free of linkage disequilibrium (LD) with each other, SNPs were further pruned based on LD and sequencing depth: within a window size of 200, each step of 20, and $r^2$ threshold of 0.1, only one of a pair of two markers that had a higher sequencing depth was retained. After pruning, 316,761 markers were retained for local ancestry analyses. The CEU and TSI datasets from the 1000 Genome Project [16] were used as European population references; the YRI and LWK datasets were used as African population references. Global ancestry was inferred as principal components using LASER v2.04 (Locating Ancestry from SEquence Reads) (S2 Fig) [17]. The overall percentage of European and African ancestry for each individual was also estimated by averaging local ancestry across the individual's entire exome. Correlations were calculated between European, African ancestral percentage and principal components estimated by LASER (S3 Fig).

For admixture mapping, the following regression equation was used:

$$Medically\ adjusted\ BP = \beta_0 + \beta_1\ G_{AA} + \beta_2\ (G_{AA} - L_{AA}) + \beta_3\ age + \beta_4\ age^2 + \beta_5\ sex + \beta_6\ BMI + \varepsilon$$

$G_{AA}$ represents global African ancestry, and $L_{AA}$ represents local African ancestry at a specific locus. By definition, global ancestry and local ancestry are correlated. Therefore, we used the difference between global and local ancestry instead of just local ancestry for the purpose of easier interpretation of the resulting $\beta_2$. $\beta_2$ is interpreted as the local response accounting for the global ancestral effect, which every locus carries, and it predicts how the additional ancestry at each particular locus would contribute to the phenotype.

To estimate the effective number of tests, we used the method described by Shriner et al. [18]. The R package "coda" [19] was used for the estimation. More specifically, an autoregressive (AR) model was fitted to the vector of local African ancestry and the spectral density at

frequency zero was evaluated. The order of the fitted AR model is chosen by minimizing the Akaike Information Criterion (AIC). The effective number of tests for chromosomes for each individual was summed and then averaged across all individuals. In ClinSeq[®], the total effective number of tests was equivalent to 66.84. Therefore, the genome-wide significant level $\alpha$ was 0.05/66.84 = 7.48E-04 (-$\log_{10} \alpha$ = 3.13), and significant regions were identified as contiguous regions within 1 unit drop of the peak LOD score, which yields approximately a 95% confidence interval.

## Regional association study

To identify SNPs that accounted for association signals discovered by admixture mapping, regional association studies were performed within admixture mapping significant regions. Since associated regions show a significant correlation between local ancestry and phenotype, we expect SNPs that drive the association signal to show a substantial allele frequency difference in different ancestral populations. Therefore, we only tested SNPs with allele frequency differences over 0.4 ($\delta > 0.4$) between European and African ancestral populations from the 1000 Genome Project. PLINK [20] was used to test the association between medically adjusted BP and genotypes at these SNPs, adjusting for the first 10 principal components to account for population structure. The following regression equation was used to perform the association analysis:

$$\text{Medically adjusted BP}$$
$$= \beta_0 + \beta_1 \text{ genotype} + \beta_2 \text{ age} + \beta_3 \text{ age}^2 + \beta_4 \text{ sex} + \beta_5 \text{ BMI} + \sum_{i=1}^{10} \beta_{i+5} \text{PC}_i + \varepsilon$$

## Conditional admixture mapping

We performed conditional admixture mapping in order to estimate the degree to which associated SNPs explained the observed admixture mapping signals. For this step, the genotypes of significant SNPs from the regional association study were included as covariates in the admixture mapping equation to test whether one or more SNPs were able to account for admixture mapping signals. $P$-values for the local ancestry coefficient $\beta_2$ were recorded to investigate if they remained significant after adjusting for those SNPs. If $\beta_2$ was not significant after inclusion of a SNP, we interpreted it as the included SNPs were able to account for local ancestry effects observed in admixture mapping.

## Replication cohort description

We attempted to replicate our findings in the Africa America Diabetes Mellitus (AADM) cohort [21]. AADM is a study of type 2 diabetes in sub-Saharan Africans. The study is comprised of 5,231 participants recruited from university medical centers in Accra and Kumasi in Ghana; Enugu, Ibadan, and Lagos in Nigeria; and Eldoret in Kenya. Blood pressure was measured in the sitting position using an oscillometric device (Omron Healthcare, Kyoto, Japan). Three readings were taken with a ten-minute interval between readings. The reported DBP and SBP values were the average of the second and third readings. Weight was measured in light clothes on an electronic scale to the nearest 0.1 kg. Height was measured with a stadiometer to the nearest 0.1 cm. Body mass index was calculated as weight (kg) divided by the square of height ($m^2$). Individuals taking antihypertensive medication were excluded, leaving 2,957 individuals for analysis. For both DBP and SBP, values were inverse-normalized after adjusting for sex, age, and age$^2$. Genotyping was performed using the Affymetrix[®] Axiom[®] Genome-Wide PanAFR Array Set ($n$ = 1,808) and the Illumina Infinium MEGA BeadChip, versions 1 ($n$ = 3,046) and 2 ($n$ = 377). For each array, quality control was performed as

described previously [22]. After excluding SNPs with a minor allele frequency < 5%, a geno-typing call rate < 90%, or a Hardy-Weinberg $P$-value < 0.001, principal components analysis was performed on 124,266 SNPs common to all three genotyping arrays. For each array, imputation was performed using the African Genome Resources reference panel available from the Sanger Imputation Service, using EAGLE2 [23] for pre-phasing and PBWT [24] for imputation. Association testing was performed using a linear mixed model in EPACTS (https://github.com/statgen/EPACTS) [25], with body mass index and the first three principal components as fixed effects and the genetic relatedness matrix as a random effect. The reason for only adjusting three principal components is that according to the Tracy-Widom test [26], only three principal components were significant. The first PC separated Kenyans from Ghanaians and Nigerians and also separated the Kenyans. The second PC separated Ghanaians from Nigerians. The third PC separated 11 Yoruba. Additional PCs did not explain significant amounts of variance (S4 Fig). To account for the fact that AADM is enriched for cases of type 2 diabetes, we included type 2 diabetes status as a covariate in the association analysis. Ancestry proportions were inferred by projecting genotype data onto a previously described reference panel [27] using ADMIXTURE version 1.3.0 [28]. This study was approved by the Institutional Review Boards at each study site, Howard University, and the National Institutes of Health and informed consent was obtained from each participant.

## Results

### ClinSeq® cohort characteristics

484 individuals passed QC and were included in the final admixture mapping analyses. The cohort characteristics are shown in Table 1. The correlation between the average African ancestry and DBP, $\log_{10}$(SBP), and MAP were 0.121, 0.064 and 0.108, respectively (*P-value* = 0.007, 0.159 and 0.017, respectively), indicating that all three blood pressure phenotypes increased as the percentage of African ancestry increased.

### Local ancestry and global ancestry inference

The average estimated African ancestry in our study sample was 76.7±12.8%. The average number of ancestral switch points per individual was 127. European ancestry percentage was highly correlated with PC1, and African ancestry percentage was highly correlated with PC2, both with a correlation coefficient of r < -0.97 (S3 Fig).

### Admixture mapping

We next performed admixture mapping to identify regions where local African ancestry was significantly associated with blood pressure phenotypes. For DBP, four regions reached

**Table 1. Data description.**

| Characteristics | ClinSeq® | AADM |
|---|---|---|
| Age (SD) | 56 (6) | 50 (13) |
| Sex (Female %) | 74% | 63% |
| BMI (SD) | 32 (10) | 27 (6) |
| Diabetes | 17% | 50% |
| Anti-Hypertension Meds | 45% | 36% |
| Median African Ancestry (IQR) | 79% (16%) | 92% (13%) |
| Mean SBP (SD) | 125 (14) | 137 (24) |
| Mean DBP (SD) | 73 (9) | 82 (13) |

exome-wide significance: 1q21.2–21.3 (β = 5.28, $P$-value = 6.39E-04), 4p15.1 (β = -7.94, $P$-value = 2.07E-04), 19q12 (β = -6.82, $P$-value = 6.56E-05), and 20p13 (β = 5.89, $P$-value = 5.04E-04) (Fig 1A and Table 2). Among these four regions, two overlapped regions where local African ancestry was significantly associated with MAP: 1q21.2–21.3 (β = 5.86, $P$-value = 5.32E-04) and 19q12 (β = -6.40, $P$-value = 6.37E-04) (Fig 1B and Table 2). For SBP, no region reached exome-wide significance.

## Regional association test

Since SNPs with large allele frequency differences between ancestral populations carry the most information about ancestry, we identified SNPs that had over 40% allele frequency differences between European and African ancestral populations from the 1000 Genome Project within significantly associated regions. For DBP, 21 SNPs had δ > 0.4 and for MAP, 14 SNPs had δ > 0.4 (Table 2). These SNPs were tested for associations with DBP and MAP, respectively. Multiple testing thresholds were calculated using a Bonferroni correction. For DBP, the significance threshold was 0.05/21 = 0.00238 and for MAP, it was 0.05/14 = 0.00357. For DBP, SNP rs4815428, which is a 3' UTR variant downstream of the gene *TMC2*, reached significance after Bonferroni correction. For MAP, four SNPs reached significance after Bonferroni correction: rs3126067, rs2184953, rs58001094, and rs771205. SNPs rs3126067, rs2184953 and rs58001094 are in high linkage disequilibrium with each other and are all coding variants in *FLG*. SNP rs771205 is a missense variant in *MINDY1* (Table 3).

## Conditional admixture mapping test

After performing the regional association study, we tested whether significant SNPs can explain admixture mapping signals by including SNP genotypes as covariates in the admixture mapping model. If local African ancestry in the admixture mapping model fails to reach significance after adjusting for SNP genotypes, it indicates that the included SNPs can explain the admixture mapping signal. For DBP, only SNP rs4815428 was significant and passed the Bonferroni correction. By including rs4815428 as a covariate in the admixture mapping model, the peak $P$-value at chr20:2187986 increased from 5.04E-04 to 0.025. Local African ancestry failed to reach exome-wide significance. At the $P$-value < 0.05 level, the peak was still significant, indicating that there might be other markers weakly contributing to the admixture mapping signal (Fig 2A). For MAP, inclusion of rs3126067 increased the peak $P$-value from 5.32E-04 to 0.022 (Fig 2B). After adjusting for both rs3126067 and rs771205, the peak $P$-value increased to 0.146 (Fig 2B). This means the admixture mapping signal was completely explained by SNPs rs3126067 and rs771205 (or variants tagged by those two SNPs).

## Replication analyses

We attempted to replicate the association of five SNPs with blood pressure phenotypes in the AADM dataset of sub-Saharan African samples. No SNPs passed the genome-wide significant threshold used in AADM (S1 Table). We also performed an association study in the four identified admixture mapping peak regions (Table 4). In chromosome 4, the top SNP rs145765242 maps to gene *PCDH7*, which is most highly expressed in aorta according to GTEx [29]. The top SNP rs145765242 has an alternative allele frequency of 0.6% in the African samples in the Genome Aggregation Database (gnomAD) [30].
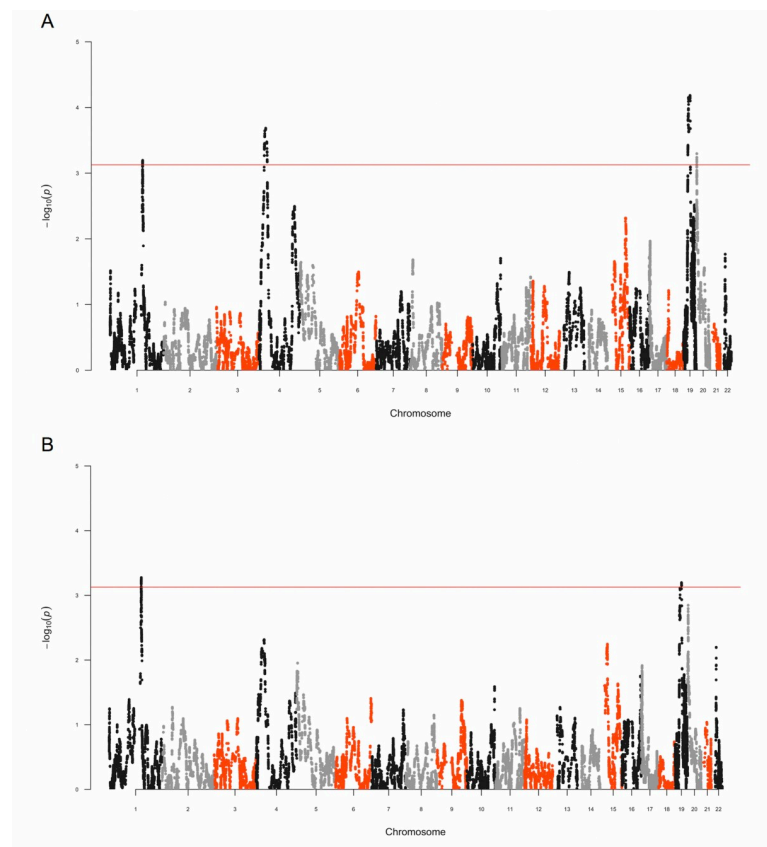
**Fig 1. Manhattan plot indicates chromosomal regions where local African ancestry is associated with DBP or MAP.** A) At chromosomal regions 1q21.2–21.3 (β = 5.28, *P*-value = 6.39E-04), 4p15.1 (β = -7.94, *P*-value = 2.07E-04), 19q12 (β = -6.82, *P*-value = 6.56E-05), and 20p13 (β = 5.89, *P*-value = 5.04E-04), local African ancestry was significantly associated with DBP. B) Two of the above regions overlapped regions where local African ancestry was significantly associated with MAP: 1q21.2–21.3 (β = 5.86, *P*-value = 5.32E-04) and 19q12 (β = -6.40, *P*-value = 6.37E-04). The red bar indicates the exome-wide significance threshold of 3.13.

https://doi.org/10.1371/journal.pone.0232048.g001

## Discussion

We utilized admixture mapping methods to identify genetic regions associated with blood pressure phenotypes in African Americans. We identified four regions for diastolic blood pressure and two regions for mean arterial pressure that reached exome-wide significance in our admixture mapping study. Two MAP regions overlapped with the DBP regions, consistent with the fact that MAP is defined partially as a function of DBP.

**Table 2. Significant regions showing the association between blood pressure and African ancestry.**

| Chr | Region | Region (Mb) | Top signal location | Beta | Std. error | *P*-value | SNPs remained after QC | SNPs with δ > 0.4 |
|-----|--------|-------------|---------------------|------|-----------|-----------|------------------------|--------------------|
| DBP | | | | | | | | |
| 1 | 1q21.2–21.3 | 148.2–154.7 | 1:151499346 | 5.28 | 1.54 | 6.39E-04 | 1034 | 16 |
| 4 | 4p15.1 | 30.7–31.1 | 4:31144153 | -7.94 | 2.12 | 2.07E-04 | 17 | 0 |
| 19 | 19q12 | 29.7–31.0 | 19:30020063 | -6.82 | 1.69 | 6.56E-05 | 84 | 0 |
| 20 | 20p13 | 0.6–5.3 | 20:2187986 | 5.89 | 1.68 | 5.04E-04 | 636 | 5 |
| MAP | | | | | | | | |
| 1 | 1q21.2–21.3 | 148.2–153.3 | 1:151501906 | 5.86 | 1.68 | 5.32E-04 | 782 | 13 |
| 19 | 19q12 | 29.7–32.1 | 19:30101339 | -6.40 | 1.86 | 6.37E-04 | 101 | 1 |

https://doi.org/10.1371/journal.pone.0232048.t002

**Table 3. Significant SNPs within admixture mapping peak regions.**

| Chr | BP | rsID | Nearby genes | Std. error | Beta | A1 | A2 | A1 frequency | | δ | *P*-value[a] | Mutation type |
|-----|-----|------|--------------|-----------|------|-----|-----|-----|-----|-----|---------|---------------|
| | | | | | | | | EUR | AFR | | | |
| DBP | | | | | | | | | | | | |
| 20 | 2597978 | rs4815428 | *TMC2* | 0.73 | -2.42 | G | A | 0.34 | 0.85 | 0.51 | 9.57E-04 | Non-coding transcript exon |
| MAP | | | | | | | | | | | | |
| 1 | 150975108 | rs771205 | *MINDY1* | 0.67 | -1.99 | T | C | 0.03 | 0.68 | 0.65 | 3.37E-03 | Missense |
| 1 | 152276889 | **rs3126067** | *FLG* | 0.74 | -2.37 | G | A | 0.15 | 0.79 | 0.64 | 1.39E-03 | Synonymous |
| 1 | 152280782 | **rs2184953** | *FLG* | 0.73 | -2.33 | G | A | 0.18 | 0.79 | 0.61 | 1.44E-03 | Missense |
| 1 | 152283862 | **rs58001094** | *FLG* | 0.73 | -2.33 | C | G | 0.18 | 0.78 | 0.60 | 1.44E-03 | Missense |

[a] Multiple testing thresholds were calculated as 0.05/21 = 2.38E-03 for DBP and 0.05/14 = 3.57E-03 for MAP. Bold SNPs indicate they are in high linkage disequilibrium with each other ($r^2$ > 0.99, D' = 1, estimated in the dataset).

https://doi.org/10.1371/journal.pone.0232048.t003

The significant admixture mapping region on chromosome 4 predominantly mapped to protocadherin 7 (*PCDH7*), which is most highly expressed in aorta according to GTEx [29]. Meta-analysis of nearly 35,000 individuals with African ancestry found SNP rs11931572, which tagged *PCDH7*, to be significantly associated with DBP [31]. The SNP rs11931572 has a low alternative allele frequency of 5% in Africans according to 1000 Genomes [16] and gnomAD [30]. This is consistent with our finding that this genetic region was significantly associated with African ancestry in both ClinSeq® and AADM datasets. However, utilizing admixture mapping drastically decreased the required sample size to identify this gene.

Within admixture mapping significant regions, five SNPs that had large ancestral allele frequency deviations (δ > 0.4) were significantly associated with BP. For DBP, SNP rs4815428 partially explained the admixture signal on chromosome 20. For MAP, SNPs rs3126067, rs2184953 and rs58001094 were in high LD with each other, and along with SNP rs771205, fully explained the admixture signal on chromosome 1.

SNPs rs3126067, rs2184953 and rs58001094 had previously been reported to be associated with ichthyosis and atopy [32] and atopic dermatitis [33]. All three SNPs are in the coding region of the gene filaggrin (*FLG)*, an intermediate filament-associated protein that aggregates keratin intermediate filaments in mammalian epidermis. Previously, this gene had been reported to be associated with asthma [34], ichthyosis [32] and abnormal inflammatory response [35, 36]. In a study of rheumatoid arthritis patients, first-degree relatives who were negative for rheumatoid arthritis but were positive for antibodies to citrullinated filaggrin had higher SBP and DBP than those who were antibody-negative [37], suggesting that there may be a correlation, although not necessarily a causation, between blood pressure and *FLG*.

SNP rs771205 is in the coding region of *MINDY1*, which encodes a hydrolase that removes lysine-48-linked conjugated ubiquitin from proteins [38]. It has exodeubiquitinase activity with a preference for long polyubiquitin chains and may play a regulatory role at the level of protein turnover [39]. This genetic region was previously reported to be linked to late-onset Alzheimer's disease [40, 41].

SNP rs4815428 is located in *TMC2*, which encodes transmembrane channel-like protein 2. *TMC2* is a potential ion channel required for the mechano-transduction of cochlear hair cells [42]. A study in the Han Chinese population reported that *TMC2* was among the top genes associated with BP response to the cold pressor test (CPT), which is associated with an increased risk of cardiovascular disease [43].

Previous admixture mapping studies of BP in African Americans were mostly performed on genotyping chip data with imputation, without much emphasis on whole exome
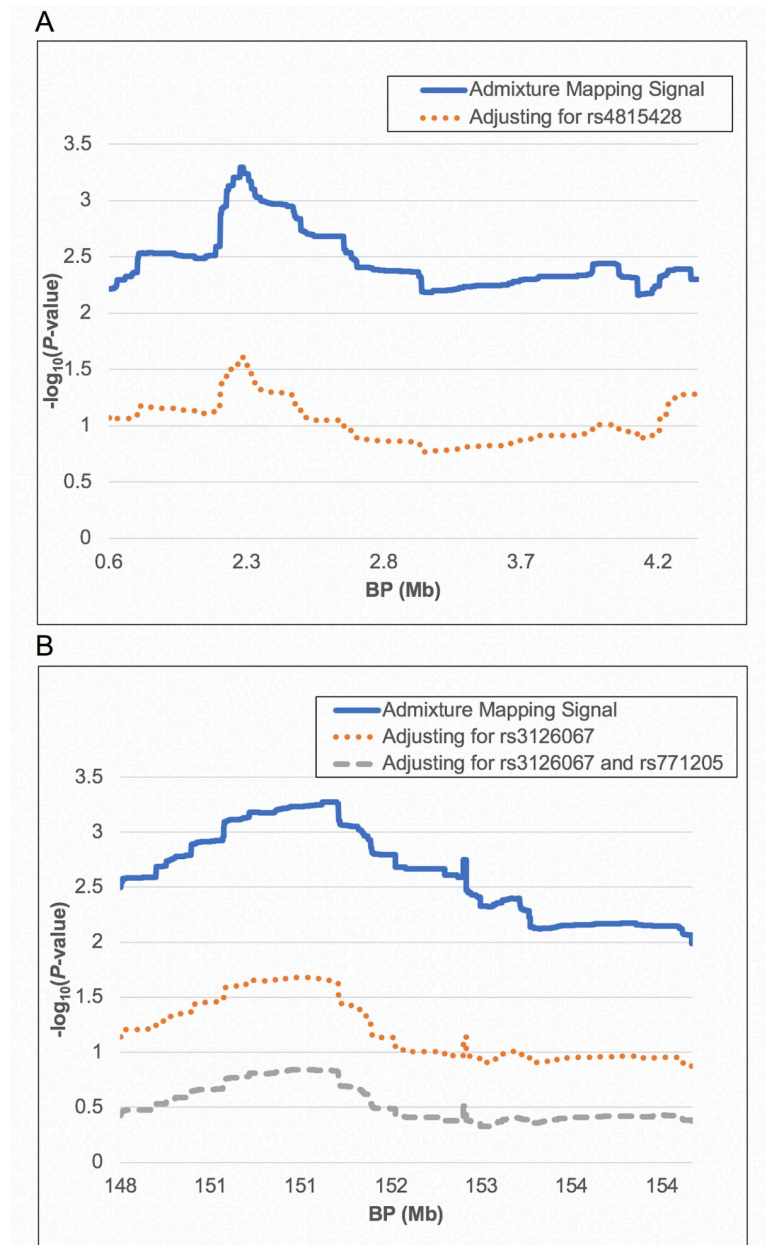
**Fig 2. Conditional admixture mapping study in DBP significant region chr20:0.6–5.3 Mb and MAP significant region chr1:148.2–153.3 Mb.** A) For DBP in the region of chr20:0.6–5.3 Mb, after adjusted for SNP rs4815428, the significant signal partially disappeared. SNP rs4815428 partially explained the significant admixture mapping signal in this region. B) For MAP in the region of chr1:148.2–153.3 Mb, after adjusted for SNPs rs3126067 and rs771205, the significant signal completely disappeared. These two SNPs fully explained the significant admixture mapping signal in this region. The x-axis is not to scale due to missing intronic regions.

https://doi.org/10.1371/journal.pone.0232048.g002

sequencing (WES) data. WES data provide greater coverage of rarer variants at exonic regions than chip data, therefore enabling us to test more variants for local ancestry association signals. Although the five SNPs we identified were all common SNPs, it is also possible for rare variants to drive the admixture mapping signals as well. With 484 African American samples, we could not identify any rare variants that passed multiple testing correction (results not shown). In

**Table 4. Top associated SNPs within admixture mapping peak regions for DBP in the AADM dataset.**

| Chr | BP | REF | ALT | rsID | Nearby gene | Alt. freq | Stat | P-value | Beta | Std. error | $R^2$ | Alt. freq$_{AFR}$ | Alt. freq$_{EUR}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 150579702 | TA | T | rs11360645 | *ENSA* | 0.198 | -3.63 | 2.84E-04 | -0.12 | 0.03 | 0.0045 | 0.192 | 0.549 |
| 4 | 30813550 | C | A | rs145765242 | *PCDH7* | 0.00947 | 3.77 | 1.65E-04 | 0.48 | 0.13 | 0.0048 | 0.00794 | 0 |
| 19 | 30950681 | C | T | rs150339768 | *ZNF536* | 0.00118 | -4.01 | 6.27E-05 | -1.45 | 0.36 | 0.0054 | 0.00099 | 0.00398 |
| 20 | 3200824 | CT | C | NA | *ITPA* | 0.425 | -4.10 | 4.31E-05 | -0.10 | 0.03 | 0.0057 | NA | NA |

https://doi.org/10.1371/journal.pone.0232048.t004

addition, utilizing WES data may leave out ancestral switch points occurring within non-exonic regions, causing the average number of switch points to be lower than those estimated using WGS or chip data.

Our five significantly-associated SNPs were not replicable in the AADM study, possibly due to the following reasons. 1) The African American individuals recruited in the Clin-Seq® study were relatively healthy individuals, mostly without an identifiable disease trait such as diabetes or cardiovascular disease, whereas AADM was enriched for cases of type 2 diabetes. However, the inclusion of type 2 diabetes status as a covariate made negligible differences in the association statistics. 2) The African ancestry in the African Americans Clin-Seq® study might not be sufficiently matched to the Africans in the AADM study. Participants were recruited for the AADM study from three countries: Nigeria, Ghana, and Kenya. Individuals from Nigeria and Ghana are expected to share ancestry with African Americans, based on historical records of the trans-Atlantic slave trade. However, the contribution to African Americans from other places such as Senegal is not as well captured by AADM. 3) The AADM participants were recruited on the African continent, while the Clin-Seq® participants were recruited in North America. Differences in environmental factors between the studies may have contributed to relatively smaller effect sizes in AADM than in ClinSeq®. 4) The causal variants may be specific to European ancestry. 5) The SNPs associated in ClinSeq® are not the SNPs underlying differential risk in the regions identified by admixture mapping.

To demonstrate that these five SNPs are not just ancestry-informative markers (i.e., confounders), we stratified samples based on their local ancestry at each SNP and performed genotypic association within each stratum, which, by definition, cannot be confounded by local ancestry. We then combined all strata and performed a random effects meta-analysis to get an overall estimation of the association for each SNP. We were able to demonstrate that genotypic association at rs4815428 was not confounded by local ancestry. This SNP has the largest effect size among all five SNPs. We were not able to demonstrate that the rest of the four SNPs were not confounded (S2 Table). After stratification, each stratum has a very small sample size. We believe that increasing the study sample size may give a more definitive answer.

In summary, we performed admixture mapping analyses on the ClinSeq® African American cohort and identified four genetic regions associated with blood pressure phenotypes. We fine-mapped these regions and identified five SNPs that are the main driving forces of associations between local African ancestry and blood pressure phenotypes at two of the four regions. SNP rs3126067, rs2184953 and rs58001094, all located in *FLG*, and SNP rs771205, located in *MINDY1*, were significantly associated with MAP. SNP rs4815428, located in *TMC2*, was significantly associated with DBP. Finally, region 4p15.1, despite containing no SNPs exhibiting large allele frequency deviation in our dataset, maps predominantly to *PCDH7*, which is most highly expressed in aorta.

## Supporting information

**S1 Fig. Distribution of systolic blood pressure, diastolic blood pressure, and mean arterial pressure for the ClinSeq® A2 dataset.** The phenotype distributions are approximately normal.
(TIFF)

**S2 Fig. Scatter plot for ClinSeq® A2 dataset in HGDP panel.** This figure shows principal components 1 and 2 of the ClinSeq® A2 dataset using the HGDP reference panel. The red cluster represents African ancestral populations; the blue cluster represents European ancestral populations; the yellow cluster represents East Asian populations, and the purple cluster represents Native American populations from the HGDP reference panel. The black cluster represents the ClinSeq® study.
(TIFF)

**S3 Fig. Correlation between global European ancestry and principal component 1 and global African ancestry and principal component 2.** This figure shows the global European ancestry estimated by averaging local ancestry across the entire exome plotted against principal component 1 as estimated by LASER, and the global African ancestry plotted against principal component 2. The X axis denotes global ancestry; the Y axis denotes principal components. Each dot denotes an African American individual who has passed QC.
(TIFF)

**S4 Fig. Principal component analysis plots of 5,231 participants from the AADM Study.** Principal component 1 separates Kenyans from Ghanaians and Nigerians and also separates the Kenyans. Principal component 2 separates Ghanaians from Nigerians. Principal component 3 separates 11 Yoruba.
(TIFF)

**S1 Table. Replication study in the AADM dataset.**
(DOCX)

**S2 Table. Meta-analyses of local-ancestry stratified genotypic association tests.**
(DOCX)

## Acknowledgments

## Author Contributions

**Conceptualization:** Zhi Liu.

**Formal analysis:** Zhi Liu.

**Funding acquisition:** James C. Mullikin.

**Methodology:** Daniel Shriner.

**Supervision:** Nancy F. Hansen, Charles N. Rotimi, James C. Mullikin.

**Validation:** Daniel Shriner.

**Writing – original draft:** Zhi Liu.

**Writing – review & editing:** Daniel Shriner, Nancy F. Hansen, Charles N. Rotimi, James C. Mullikin.

## References

1. Ehret GB, Caulfield MJ. Genes for blood pressure: an opportunity to understand hypertension. Eur Heart J. 2013; 34(13):951–61. Epub 2013/01/11. https://doi.org/10.1093/eurheartj/ehs455 PMID: 23303660; PubMed Central PMCID: PMC3612776.

2. Ehret GB, Bakris B, Forman J. Genetic factors in the pathogenesis of hypertension. UpToDate. 2017; Topic 3847 Version 17.0.

3. Fryar CD, Ostchega Y, Hales CM, Zhang G, Kruszon-Moran D. Hypertension Prevalence and Control Among Adults: United States, 2015–2016 NCHS Data Brief No. 2892017. Available from: https://www.cdc.gov/nchs/products/databriefs/db289.htm.

4. Kramer H, Han C, Post W, Goff D, Diez-Roux A, Cooper R, et al. Racial/ethnic differences in hypertension and hypertension treatment and control in the multi-ethnic study of atherosclerosis (MESA). Am J Hypertens. 2004; 17(10):963–70. https://doi.org/10.1016/j.amjhyper.2004.06.001 PMID: 15485761.

5. Maass PG, Aydin A, Luft FC, Schachterle C, Weise A, Stricker S, et al. PDE3A mutations cause autosomal dominant hypertension with brachydactyly. Nat Genet. 2015; 47(6):647–53. Epub 2015/05/12. https://doi.org/10.1038/ng.3302 PMID: 25961942.

6. Salvi E, Kutalik Z, Glorioso N, Benaglio P, Frau F, Kuznetsova T, et al. Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. Hypertension. 2012; 59(2):248–55. Epub 2011/12/21. https://doi.org/10.1161/HYPERTENSIONAHA.111.181990 PMID: 22184326; PubMed Central PMCID: PMC3272453.

7. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature genetics. 2011; 43 (4):333–8. https://doi.org/10.1038/ng.784 PMID: 21378990

8. Zhu X, Luke A, Cooper RS, Quertermous T, Hanis C, Mosley T, et al. Admixture mapping for hypertension loci with genome-scan markers. Nat Genet. 2005; 37(2):177–81. Epub 2005/01/25. https://doi.org/10.1038/ng1510 PMID: 15665825.

9. Shetty PB, Tang H, Tayo BO, Morrison AC, Hanis CL, Rao DC, et al. Variants in CXADR and F2RL1 are associated with blood pressure and obesity in African-Americans in regions identified through admixture mapping. J Hypertens. 2012; 30(10):1970–6. Epub 2012/08/24. https://doi.org/10.1097/HJH.0b013e3283578c80 PMID: 22914544; PubMed Central PMCID: PMC3575678.

10. Zhu X, Young JH, Fox E, Keating BJ, Franceschini N, Kang S, et al. Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: contributions from the CARe consortium. Human molecular genetics. 2011; 20(11):2285–95. https://doi.org/10.1093/hmg/ddr113 PMID: 21422096

11. Sofer T, Baier LJ, Browning SR, Thornton TA, Talavera GA, Wassertheil-Smoller S, et al. Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits. PLoS One. 2017; 12(11):e0188400. Epub 2017/11/21. https://doi.org/10.1371/journal.pone.0188400 PMID: 29155883; PubMed Central PMCID: PMC5695820.

12. Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome Res. 2010; 20 (10):1420–31. Epub 2010/09/03. https://doi.org/10.1101/gr.106716.110 PMID: 20810667; PubMed Central PMCID: PMC2945191.

13. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76(5):887–93. Epub 2005/03/25. https://doi.org/10.1086/429864 PMID: 15789306; PubMed Central PMCID: PMC1199378.

14. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009; 41(6):677–87. Epub 2009/05/12. https://doi.org/10.1038/ng.384 PMID: 19430479; PubMed Central PMCID: PMC2998712.

15. Hu Y, Willer C, Zhan X, Kang HM, Abecasis GR. Accurate local-ancestry inference in exome-sequenced admixed individuals via off-target sequence reads. Am J Hum Genet. 2013; 93(5):891–9. https://doi.org/10.1016/j.ajhg.2013.10.008 PMID: 24210252; PubMed Central PMCID: PMC3824118.

16. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015; 526(7571):68–74. https://doi.org/10.1038/nature15393 PMID: 26432245; PubMed Central PMCID: PMC4750478.

17. Wang C, Zhan X, Liang L, Abecasis GR, Lin X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. Am J Hum Genet. 2015; 96(6):926–37. Epub 2015/06/02. https://doi.org/10.1016/j.ajhg.2015.04.018 PMID: 26027497; PubMed Central PMCID: PMC4457959.

18. Shriner D, Adeyemo A, Rotimi CN. Joint ancestry and association testing in admixed individuals. PLoS Comput Biol. 2011; 7(12):e1002325. Epub 2012/01/05. https://doi.org/10.1371/journal.pcbi.1002325 PMID: 22216000; PubMed Central PMCID: PMC3245293.

19. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006; 6:7–11.

20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–75. https://doi.org/10.1086/519795 PMID: 17701901; PubMed Central PMCID: PMC1950838.

21. Rotimi CN, Dunston GM, Berg K, Akinsete O, Amoah A, Owusu S, et al. In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. Ann Epidemiol. 2001; 11(1):51–8. Epub 2001/02/13. https://doi.org/10.1016/s1047-2797(00)00180-0 PMID: 11164120.

22. Adeyemo AA, Tekola-Ayele F, Doumatey AP, Bentley AR, Chen G, Huang H, et al. Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans. Front Genet. 2015; 6:335. Epub 2015/12/05. https://doi.org/10.3389/fgene.2015.00335 PMID: 26635871; PubMed Central PMCID: PMC4656823.

23. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016; 48(11):1443–8. Epub 2016/10/28. https://doi.org/10.1038/ng.3679 PMID: 27694958; PubMed Central PMCID: PMC5096458.

24. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. 2014; 30(9):1266–72. Epub 2014/01/15. https://doi.org/10.1093/bioinformatics/btu014 PMID: 24413527; PubMed Central PMCID: PMC3998136.

25. Kang HM. EPACTS: Efficient and Parallelizable Association Container Toolbox, version 3.2.6. https://github.com/statgen/EPACTS.

26. Shriner D. Improved eigenanalysis of discrete subpopulations and admixture using the minimum average partial test. Hum Hered. 2012; 73(2):73–83. Epub 2012/03/24. https://doi.org/10.1159/000335899 PMID: 22441298; PubMed Central PMCID: PMC3370671.

27. Baker JL, Rotimi CN, Shriner D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. Sci Rep. 2017; 7(1):1572. Epub 2017/05/10. https://doi.org/10.1038/s41598-017-01837-7 PMID: 28484253; PubMed Central PMCID: PMC5431528.

28. Shringarpure SS, Bustamante CD, Lange K, Alexander DH. Efficient analysis of large datasets and sex bias with ADMIXTURE. BMC Bioinformatics. 2016; 17:218. Epub 2016/05/25. https://doi.org/10.1186/s12859-016-1082-x PMID: 27216439; PubMed Central PMCID: PMC4877806.

29. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013; 45(6):580–5. Epub 2013/05/30. https://doi.org/10.1038/ng.2653 PMID: 23715323; PubMed Central PMCID: PMC4010069.

30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. 2019:531210. https://doi.org/10.1101/531210%JbioRxiv

31. Sung YJ, Winkler TW, de Las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. Am J Hum Genet. 2018; 102(3):375–400. Epub 2018/02/20. https://doi.org/10.1016/j.ajhg.2018.01.015 PMID: 29455858; PubMed Central PMCID: PMC5985266.

32. Kiritsi D, Valari M, Fortugno P, Hausser I, Lykopoulou L, Zambruno G, et al. Whole-exome sequencing in patients with ichthyosis reveals modifiers associated with increased IgE levels and allergic sensitizations. J Allergy Clin Immunol. 2015; 135(1):280–3. Epub 2014/12/03. https://doi.org/10.1016/j.jaci.2014.09.042 PMID: 25458912.

33. Kim M, Yoo J, Kim J, Park J, Han E, Jang W, et al. Association of FLG single nucleotide variations with clinical phenotypes of atopic dermatitis. PLoS One. 2017; 12(12):e0190077. Epub 2017/12/28. https://doi.org/10.1371/journal.pone.0190077 PMID: 29281699; PubMed Central PMCID: PMC5744964.

34. Rice NE, Patel BD, Lang IA, Kumari M, Frayling TM, Murray A, et al. Filaggrin gene mutations are associated with asthma and eczema in later life. J Allergy Clin Immunol. 2008; 122(4):834–6. Epub 2008/09/02. https://doi.org/10.1016/j.jaci.2008.07.027 PMID: 18760831; PubMed Central PMCID: PMC2775129.

35. Barnes KC. An update on the genetics of atopic dermatitis: scratching the surface in 2009. J Allergy Clin Immunol. 2010; 125(1):16–29 e1-11; quiz 30–1. Epub 2010/01/30. https://doi.org/10.1016/j.jaci.2009.11.008 PMID: 20109730; PubMed Central PMCID: PMC2874322.

36. Boguniewicz M, Leung DY. Atopic dermatitis: a disease of altered skin barrier and immune dysregulation. Immunol Rev. 2011; 242(1):233–46. Epub 2011/06/21. https://doi.org/10.1111/j.1600-065X.2011.01027.x PMID: 21682749; PubMed Central PMCID: PMC3122139.

37. Hughes-Austin JM, Gan RW, Deane KD, Weisman MH, Demoruelle MK, Sokolove J, et al. Association of Antibodies to Citrullinated Protein Antigens with Blood Pressure in First-Degree Relatives of Rheumatoid Arthritis Patients: The Studies of the Etiology of Rheumatoid Arthritis. Am J Nephrol. 2017; 46(6):481–7. Epub 2017/12/14. https://doi.org/10.1159/000485259 PMID: 29237149; PubMed Central PMCID: PMC5788181.

38. Kristariyanto YA, Abdul Rehman SA, Weidlich S, Knebel A, Kulathu Y. A single MIU motif of MINDY-1 recognizes K48-linked polyubiquitin chains. EMBO Rep. 2017; 18(3):392–402. Epub 2017/01/14. https://doi.org/10.15252/embr.201643205 PMID: 28082312; PubMed Central PMCID: PMC5331195.

39. Abdul Rehman SA, Kristariyanto YA, Choi SY, Nkosi PJ, Weidlich S, Labib K, et al. MINDY-1 Is a Member of an Evolutionarily Conserved and Structurally Distinct New Family of Deubiquitinating Enzymes. Mol Cell. 2016; 63(1):146–55. Epub 2016/06/14. https://doi.org/10.1016/j.molcel.2016.05.009 PMID: 27292798; PubMed Central PMCID: PMC4942677.

40. Liu F, Arias-Vasquez A, Sleegers K, Aulchenko YS, Kayser M, Sanchez-Juan P, et al. A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. Am J Hum Genet. 2007; 81(1):17–31. Epub 2007/06/15. https://doi.org/10.1086/518720 PMID: 17564960; PubMed Central PMCID: PMC1950931.

41. Butler AW, Ng MY, Hamshere ML, Forabosco P, Wroe R, Al-Chalabi A, et al. Meta-analysis of linkage studies for Alzheimer's disease—a web resource. Neurobiol Aging. 2009; 30(7):1037–47. Epub 2009/04/14. https://doi.org/10.1016/j.neurobiolaging.2009.03.013 PMID: 19362756.

42. Martinac B. The ion channels to cytoskeleton connection as potential mechanism of mechanosensitivity. Biochim Biophys Acta. 2014; 1838(2):682–91. Epub 2013/07/28. https://doi.org/10.1016/j.bbamem.2013.07.015 PMID: 23886913.

43. Yang X, Gu D, He J, Hixson JE, Rao DC, Lu F, et al. Genome-wide linkage and regional association study of blood pressure response to the cold pressor test in Han Chinese: the genetic epidemiology network of salt sensitivity study. Circ Cardiovasc Genet. 2014; 7(4):521–8. Epub 2014/07/17. https://doi.org/10.1161/CIRCGENETICS.113.000332 PMID: 25028485; PubMed Central PMCID: PMC4264620.